

FUR0011-US

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of:

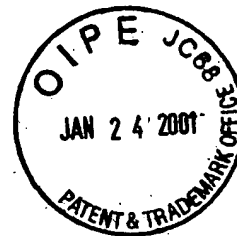
Toshio HIRAI ET AL.

Serial No. 09/678,544

Art Unit: 2641

Filed: October 4, 2000

Examiner: Unknown



For: DEVICE AND METHOD FOR
SYNTHESIZING SPEECH

Assistant Commissioner of Patents
Washington, D.C. 20231

CLAIM TO PRIORITY UNDER 35 U.S.C. § 119

Sir:

The benefit of the filing date of the following prior application filed in the following foreign country is hereby requested and the right of the priority provided under 35 U.S.C. § 119 is hereby claimed:

Japanese Patent Appln. No. 11-285125 filed October 6, 1999

In support of this claim, filed herewith is a certified copy of said foreign application.

Respectfully submitted,

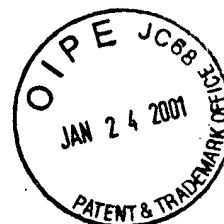
Date: January 24, 2001

SHAW PITTMAN
2300 N Street, N.W.
Washington, D.C. 20037-1128
Tel: (202) 663-8906

By:

Michael D. Bednarek
Reg. No. 32,329

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出 願 年 月 日

Date of Application:

1999年10月 6日

出 願 番 号

Application Number:

平成11年特許願第285125号

出 願 人

Applicant(s):

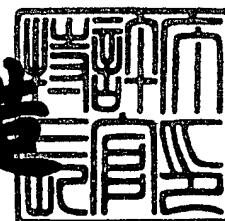
株式会社アルカディア

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年10月13日

特 許 庁 長 官
Commissioner,
Patent Office

及 川 耕 造



出証番号 出証特2000-3083596

【書類名】 特許願

【整理番号】 ARK012

【提出日】 平成11年10月 6日

【あて先】 特許庁長官 殿

【国際特許分類】 G01L 3/00

【発明者】

 【住所又は居所】 大阪府箕面市坊島 1 丁目 3 番 4 0 号 株式会社アルカデ
 ィア内

 【氏名】 天白 成一

【発明者】

 【住所又は居所】 大阪府箕面市坊島 1 丁目 3 番 4 0 号 株式会社アルカデ
 ィア内

 【氏名】 平井 俊男

【特許出願人】

 【識別番号】 593160323

 【氏名又は名称】 株式会社アルカディア

【代理人】

 【識別番号】 100092956

 【弁理士】

 【氏名又は名称】 古谷 栄男

 【電話番号】 06-6368-2160

【選任した代理人】

 【識別番号】 100101018

 【弁理士】

 【氏名又は名称】 松下 正

 【電話番号】 06-6368-2160

【選任した代理人】

 【識別番号】 100101546

 【弁理士】

【氏名又は名称】 眞島 宏明

【電話番号】 06-6368-2160

【選任した代理人】

【識別番号】 100106013

【弁理士】

【氏名又は名称】 田川 幸一

【電話番号】 06-6368-2160

【手数料の表示】

【予納台帳番号】 004891

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声合成装置および方法

【特許請求の範囲】

【請求項 1】

音声単位のサンプル波形データと、これに音韻情報を関連づけて形成した音声データベースを記録した音声データベース記録手段と、

出力すべき音声の音韻情報を受けて、この音韻情報を音声単位に区分するとともに、前記音声データベースから、音声単位に区分したそれぞれの音韻情報について対応するサンプル音声波形データを取得し、取得した音声単位のサンプル音声波形データを結合して出力すべき音声波形データを得る音声波形合成手段と、

音声波形合成手段によって得られた音声波形データを受けて、アナログ音声信号に変換するアナログ変換手段と、

を備えた音声合成装置であって、

前記音声波形合成手段は、音声波形データの周期的単位区間について、マイナスピークに向かって変化する区間の波形を加工することにより、ピッチを変更するピッチ変更手段を備えている音声合成装置。

【請求項 2】

ピッチ変更処理をコンピュータに行わせるためのプログラムを記録した記録媒体であって、

ピッチ変更すべき音声波形データを受け取り、当該音声波形データの周期的単位区間について、マイナスピークに向かって変化する区間の波形を加工することにより、ピッチ変更を施す処理をコンピュータに行わせるためのプログラムを記録した記録媒体。

【請求項 3】

各音声単位毎にピッチの異なるサンプル音声波形データを保持し、当該音声波形データに対応付けて音韻情報を保持した音声データベースを記録した音声データベース記録手段と、

出力すべき音声の音韻情報を受けて、この音韻情報を音声単位に区分するとともに、音韻情報に基づいて、前記音声データベースから、区分された音韻情報に

対応するサンプル音声波形データのうち、好ましいピッチのサンプル音声波形データを取得し、取得した音声単位のサンプル音声波形データを結合して出力すべき音声波形データを得る音声波形合成手段と、

音声波形合成手段によって得られた音声波形データを受けて、アナログ音声信号に変換するアナログ変換手段と、

を備えた音声合成装置であって、

前記音声データベースは、各音声単位毎に、音声波形データの周期的単位区間について、マイナスピークに向かって変化する区間の波形を異ならせるように加工してピッチを異ならせた複数のサンプル音声波形データを有するよう構成された音声合成装置。

【請求項 4】

音声データベースを用いて音声波形合成処理をコンピュータに行わせるためのプログラムを記録した記録媒体であって、

出力すべき音声の音韻情報を受けて、この音韻情報を音声単位に区分するとともに、音韻情報に基づいて、前記音声データベースから、区分された音韻情報に対応するサンプル音声波形データのうち、好ましいピッチのサンプル音声波形データを取得し、取得した音声単位のサンプル音声波形データを結合して出力すべき音声波形データを得る音声波形合成処理をコンピュータに行わせるためのプログラムを記録した記録媒体。

ここで、前記音声データベースは、各音声単位毎に、音声波形データの周期的単位区間について、マイナスピークに向かって変化する区間の波形を異ならせるように加工してピッチを異ならせた複数のサンプル音声波形データを有するよう構成されている。

【請求項 5】

各音声単位毎にピッチの異なる複数のサンプル音声波形データを記録した記録媒体であって、

前記複数のサンプル音声波形データは、音声波形データの周期的単位区間について、マイナスピークに向かって変化する区間の波形を異ならせるように加工することにより、ピッチを異ならせていることを特徴とするもの。

【請求項 6】

請求項 1 ～ 5 のいずれか音声合成装置または記録媒体において、

前記マイナスピークに向かって変化する区間のうち、ゼロクロス近傍における加工値を最大にし、ゼロクロスより離れるほど加工値を小さくするようにしたことを特徴とするもの。

【請求項 7】

請求項 1 ～ 6 のいずれかの音声合成装置または記録媒体において、

前記マイナスピークに向かって変化する区間において、波形を時間的に圧縮または伸長するよう加工することにより、ピッチを短くまたは長くすることを特徴とするもの。

【請求項 8】

請求項 1 ～ 5 のいずれか音声合成装置または記録媒体において、

前記マイナスピークに向かって変化する区間のうち、ゼロクロス近傍において加工を行うことを特徴とするもの。

【請求項 9】

請求項 8 の音声合成装置または記録媒体において、

前記ゼロクロス近傍において、実質的なゼロ値を挿入することにより、ピッチを長くする処理、または実質的なゼロ値部分を削除することにより、ピッチを短くする処理の、少なくとも一方の処理を行うことを特徴とするもの。

【請求項 1 0】

人間の発話に基づく音声単位のサンプル波形データを記録するサンプル波形データ記録領域と、

各音声単位のサンプル波形データに対応付けて音韻情報を記録する音韻情報記録領域と、

サンプル波形データのマイナスピークの直前のゼロクロスを示すゼロクロス指示情報を記録する指示情報記録領域と、

を備えた音声データベースを記録した記録媒体。

【請求項 1 1】

音声波形に対するピッチ変更方法であって、

音声波形の周期的単位区間について、マイナスピークに向かって変化する区間の波形を加工することによりピッチ変更を施すこと、
を特徴とするピッチ変更方法。

【請求項 1 2】

入力された指令に応じて音声波形に対する処理を行う音声処理装置であって、
上方向矢印、下方向矢印、右方向矢印、左方向矢印を示すアイコンまたはスイッチに対応付けて、少なくとも、音声の強度、ピッチ周波数または発話時間長のいずれか一つを変更する処理を行うことを特徴とする音声処理装置。

【請求項 1 3】

入力された指令に応じて音声波形に対する処理を行う音声処理装置をコンピュータによって実現するためのプログラムを記録した記録媒体であって、
上方向矢印、下方向矢印、右方向矢印、左方向矢印を示すアイコンまたはスイッチに対応付けて、少なくとも、波形の強度、ピッチ周波数または発話時間長のいずれか一つを変更する処理を、コンピュータに行わせるためのプログラムを記録した記録媒体。

【請求項 1 4】

入力された指令に応じて音声波形に対する処理を行う音声処理装置であって、
上方向矢印に対応づけて、少なくともピッチ周波数を高くする処理を行い、下方向矢印に対応づけて、少なくともピッチ周波数を低くする処理を行うことを特徴とする音声処理装置。

【請求項 1 5】

入力された指令に応じて音声波形に対する処理を行う音声処理装置をコンピュータによって実現するためのプログラムを記録した記録媒体であって、
上方向矢印に対応づけて、少なくともピッチ周波数を高くする処理を行い、下方向矢印に対応づけて、少なくともピッチ周波数を低くする処理を、コンピュータに行わせるためのプログラムを記録した記録媒体。

【発明の詳細な説明】

【0 0 0 1】

【発明の技術分野】

この発明は音声合成等の音声処理に関するものであり、特にピッチ変更処理に関するものである。

【0002】

【従来の技術および発明が解決しようとする課題】

音声合成方式の一つとして、「波形接続型」と呼ばれるものがある。これは、予め用意した波形を接続して、目的とする音声を合成するものである。しかし、波形を接続するだけでは、音声のイントネーションが制御できず、自然な合成音声を得られないという問題があった。

【0003】

これを解決する方法として、音声のピッチ2つ分をハニング窓などで切り出し、それを少しずつ重ね合わせ、ピッチ長の異なった音声を得るPSOLA (Pitch Synchronous OverLap Add) 法が提案されている (E. Moulines et. al, "Pitch-Synchronous waveform processing techniques for text-to-speech synthesis using diphones" Speech Communication, 1990.9)。

【0004】

図22、図23を用いて、PSOLA法を説明する。図22は、音声波形の一部を示したものである。図に示すように、ほぼ周期的な波形が繰り返されている。この繰り返し部分をピッチといい、ピッチの長さにより、音声の高さが変わる。

【0005】

PSOLA法は、図23に示すように、まず、ピーク位置Mを中心として、ハニング窓で切り出しを行う。次に、切り出したそれぞれの波形を目的のピッチ長となるように重ね合わせる。なお、切り出しに用いるハニング窓の幅は、重ね合わせの際に、 $1/2$ オーバーラップする幅とする。このようにすれば、余分な周波数成分をもたらすことを抑えつつ、ピッチの変更を行うことができる。上記のようにPSOLA法によって、ピッチを加工（つまり基本周波数を加工）すれば、イントネーションを制御することができる。

【0006】

しかしながら、このようなPSOLA法には、次のような問題があった。

【0007】

第一に、図 2 4 ～ 図 2 7 に示すように、ハニング窓での重ね合わせ部分において、不自然な振幅の低減がもたらされてしまうおそれがある。図 2 4 は原波形（理解を容易にするため、減衰正弦波にて示した）、図 2 5 はハニング窓の左側成分によって切り出された波形、図 2 6 はハニング窓の右側成分によって切り出された波形、図 2 7 は合成された波形である。図 2 7 に示すように、中央部分において、不自然な振幅の低減がもたらされている。これにより、ホルマントで表現される音声波形の微細構造が破壊されてしまう。

【 0 0 0 8 】

第二に、図 2 8 に示すように、隣接するピッチのピークによって、エコーが生じてしまうという問題である。これは、河井恒他「波形素片接続型音声合成システムの検討」（電子情報通信学会、音声研究会（技術資料）、1993 年 5 月）において、指摘されている。同文献では、これを解決するため、窓の裾を短かくした台形型の窓を用いることが提案されている。しかしながら、かかる台形型の窓では、重ね合わせにより、余分な周波数成分がもたらされ、合成音が不自然になるおそれがある。

【 0 0 0 9 】

この発明では、上記のような問題点を解決して、音声の自然性を損なうことの少ないピッチ変更処理技術を提供することを目的とする。

【 0 0 1 0 】

【課題を解決するための手段】

音声波形の各ピッチにおける波形は、図 1 に示すように、1) 声帯の閉鎖に伴うマイナスピークが出現する時点 α 、2) マイナスピークを起点とした声道の形状に基づく波形が現れる区間 β 、3) 声道形状に基づく音声波形が減衰しながら次のマイナスピークに向かって収束する区間 γ に分けられる。上記の PSOLA 法では、ピーク M 付近の波形形状を維持することに着目して、ピーク M 付近にハニング窓の中心を持ってきている。しかし、ピーク付近での波形形状の維持に着目するあまり、上記のような問題点が生じている。

【 0 0 1 1 】

これに対し、発明者らは、上記の音声波形の特性に基づいて、声帯の閉鎖に伴

うマイナスピークによる影響が小さい部分（つまり次のマイナスピークの直前である区間 γ ）において、ピッチ変更のための波形加工を行うことを着想した。これによれば、ピーク付近での波形形状を完全に維持しつつ、かつ、ピッチ変更の影響を最小限に抑えつつ、加工を行うことができる。

【 0 0 1 2 】

また、発明者らは、音声波形のピッチについて、どの部分が共通しているのか、あるいはどの部分が異なっているのかを、採取した音声に基づいて検討した。図 2 に、いくつかの「あ」の波形の 1 ピッチ分を示す。この図から明らかなように、2.5ms まではほぼ似ていて、それ以降ほぼゼロ値が続き、ある時点からは単調減少してマイナスのピーク値に向かっている。つまり、実際の発話においてピッチのことなる波形は、ゼロ値部分の長さの違い（単調減少区間の位置の違い）として特徴付けられることが分かった。したがって、図 1 の区間 γ のうち、特に、ゼロ値の部分加工することにより、適切なピッチ加工を行い得ることを見いだした。

【 0 0 1 3 】

請求項 1、3 の音声合成装置は、音声波形データの周期的単位区間について、マイナスピークに向かって変化する区間の波形を加工することにより、ピッチを変更するピッチ変更手段を備えている。

【 0 0 1 4 】

したがって、声帯の閉鎖に伴うマイナスピークによる影響が小さい部分において波形を加工することができ、自然性を保持したピッチ変更を行うことができる。

【 0 0 1 5 】

請求項 6 の音声合成装置は、マイナスピークに向かって変化する区間のうち、ゼロクロス近傍における加工値を最大にし、ゼロクロスより離れるほど加工値を小さくするようにしたことを特徴としている。

【 0 0 1 6 】

したがって、ゼロ値部分の長さが異なるという実際の音声の傾向に沿った波形加工を行ってピッチを調整することができ、自然性を損なわないピッチ変更を行

うことができる。

【 0 0 1 7 】

請求項 7 の音声合成装置は、前記マイナスピークに向かって変化する区間において、波形を時間的に圧縮または伸長するよう加工することにより、ピッチを短くまたは長くすることを特徴としている。

【 0 0 1 8 】

したがって、声帯の閉鎖に伴うマイナスピークによる影響が小さい部分において波形を時間的に加工することができ、自然性を保持したピッチ変更を行うことができる。

【 0 0 1 9 】

請求項 8 の音声合成装置は、マイナスピークに向かって変化する区間のうち、ゼロクロス近傍において加工を行うことを特徴としている。したがって、振幅が小さく、より影響の少ない部分において、加工を行うことができる。

【 0 0 2 0 】

請求項 9 の音声合成装置は、ゼロクロス近傍において、実質的なゼロ値を挿入することにより、ピッチを長くする処理、または実質的なゼロ値部分を削除することにより、ピッチを短くする処理の、少なくとも一方の処理を行うことを特徴としている。

【 0 0 2 1 】

したがって、スペクトルに与える影響を最小限に抑えつつ、ピッチ変更を行うことができる。また、ゼロ値の挿入、ゼロ値の削除という極めて簡易な処理であるため、処理が迅速である。

【 0 0 2 2 】

請求項 1 1 の音声波形に対するピッチ変更方法は、音声波形の周期的単位区間について、マイナスピークに向かって変化する区間の波形を加工することによりピッチ変更を施すことを特徴としている。

【 0 0 2 3 】

したがって、声帯の閉鎖に伴うマイナスピークによる影響が小さい部分において波形を加工することができ、自然性を保持したピッチ変更を行うことができる

【 0 0 2 4 】

請求項 1 2 の音声処理装置は、上方向矢印、下方向矢印、右方向矢印、左方向矢印を示すアイコンまたはスイッチに対応付けて、少なくとも、音声の強度、ピッチ周波数または発話時間長のいずれか一つを変更する処理を行うことを特徴としている。

【 0 0 2 5 】

したがって、音声の強度、ピッチ周波数または発話時間長の変更を簡単な操作で行うことができる。

【 0 0 2 6 】

請求項 1 4 の音声処理装置は、上方向矢印に対応づけて、少なくともピッチ周波数を高くする処理を行い、下方向矢印に対応づけて、少なくともピッチ周波数を低くする処理を行うことを特徴としている。

【 0 0 2 7 】

したがって、ピッチ変更処理について、直感的に処理内容を理解しやすい操作インターフェイスを提供することができる。

【 0 0 2 8 】

この発明において、「音声単位」とは、音声合成または解析の際に音声波形をひとかたまりとして扱う単位をいう。

【 0 0 2 9 】

「音声データベース」とは、少なくとも音声波形とこれに対応する音韻情報を記録したデータベースをいう。

【 0 0 3 0 】

「音声波形合成手段」とは、規則もしくはサンプル波形に基づいて、与えられた音韻情報に対応する音声波形を生成する手段をいう。実施形態においては、たとえば、図 5、図 6 のステップ S 4 ～ S 1 2 がこれに対応する。

【 0 0 3 1 】

「周期的単位区間」とは、音声波形において周期的に繰り返される 1 つの区間をいう。実施形態においては、ピッチがこれに該当する。

【 0 0 3 2 】

「矢印」とは、方向性を明示もしくは示唆する記号をいうものであり、三角形によって方向を示したもの等も含む概念である。

【 0 0 3 3 】

「プログラム（データ）を記録した記録媒体」とは、プログラム（データ）を記録した ROM、RAM、フレキシブルディスク、CD-ROM、メモリカード、ハードディスク等の記録媒体をいう。また、電話回線、搬送路等の通信媒体も含む概念である。CPU に接続されて、記録されたプログラムが直接実行されるハードディスクのような記録媒体だけでなく、一旦ハードディスク等にインストールした後に実行されるプログラムを記録した CD-ROM 等の記録媒体を含む概念である。さらに、ここでいうプログラム（データ）には、直接実行可能なプログラムだけでなく、ソース形式のプログラム、圧縮処理がされたプログラム（データ）、暗号化されたプログラム（データ）等を含む。

【 0 0 3 4 】

【発明の実施の形態】

1. 第 1 の実施形態

(1) 全体構成

図 3 に、この発明の一実施形態による音声合成装置の全体構成を示す。この実施形態では、音声波形合成手段 16 は、文字列解析手段 2、音声単位取得手段 4、波形変更手段 12、波形結合手段 22 を備えている。また、波形変更手段 12 は、継続長変更手段 6、強度変更手段 8、ピッチ変更手段 10 を備えている。

【 0 0 3 5 】

入力された文字列は、解析用辞書 20 を参照して、文字列解析手段 2 によって形態素解析がなされ、音声単位に分離される。さらに、解析用辞書 20 を参照し、前後の音の環境等を考慮して、各音声単位ごとに、無声音／有声音の区別、継続時間長、強度の時間的变化、基本周波数の時間的变化を算出する。

【 0 0 3 6 】

音声単位取得手段 4 は、文字列解析手段 2 の形態素解析の結果を受けて、音声データベース 18 から、各音声単位のサンプル音声波形を取得する。継続長変更

手段 6 は、文字列解析手段 2 から与えられた継続時間長に基づいて、取得したサンプル音声波形の継続時間長を変更する。強度変更手段 8 は、文字列解析手段 2 から与えられた強度の時間的变化に基づいて、取得したサンプル音声波形の強度を変更する。ピッチ変更手段 10 は、文字列解析手段 2 から与えられた基本周波数の時間的变化に基づいて、取得したサンプル音声波形のピッチを変更する。上記のようにして所望の変更を受けた各音声単位のサンプル音声波形は、波形結合手段 22 によって接続され、音声波形データとして出力される。

【0037】

アナログ変換手段 14 は、この音声波形データをアナログ変換して音声信号として出力する。

【0038】

(2) ハードウェア構成

図 4 に、図 3 の音声合成装置を CPU を用いて実現した場合のハードウェア構成例を示す。CPU 30 には、メモリ 32、キーボード/マウス 34、フロッピーディスクドライブ (FDD) 36、CD-ROM ドライブ 40、ハードディスク 44、アナログ変換手段であるサウンドカード 54、ディスプレイ 58 が接続されている。ハードディスク 44 には、オペレーティングシステム (OS) 52 (たとえば、マイクロソフト社の WINDOWS98 など)、音声合成プログラム 46 が格納されている。また、音声データベース 48、解析用辞書 50 も格納されている。これらプログラムは、CD-ROM ドライブ 40 を介して、CD-ROM 42 からインストールされたものである。

【0039】

この実施形態では、音声合成プログラム 46 は、OS 52 と共同してその各機能を実現している。しかし、その一部または全部を、音声合成プログラム 46 が単独で実現するようにしてもよい。

【0040】

(3) 音声合成処理

図 5 に、ハードディスク 44 に記録された音声合成プログラムのフローチャートを示す。まず、操作者からキーボード 34 により、出力すべき音声に対応する

文字列が入力される（ステップ S 1）。なお、この文字列は、フロッピーディスク 38 に記録されたものやネットワークを介して他のコンピュータから送られたものであってもよい。

【0041】

次に、CPU 30 は、この文字列について、解析用辞書 50 の単語辞書を参照して、形態素解析を行う（ステップ S 2）。この単語辞書の構成例を図 8 に示す。CPU 30 は、この単語辞書を参照し、文字列を単語に分解しつつ、その読みを得る。たとえば、「こんにちは」という文字列が入力された場合には、「koNn ichiwa」という読みを得る。

【0042】

さらに、各単語について、単語を構成する音節のアクセント値を取得する（ステップ S 3）。したがって、「ko」「N」「ni」「chi」「wa」という音節が得られ、それぞれの音節について、図 8 に示すアクセント値が得られる。アクセント値は、前後の音の環境によって変化するものである。よって、CPU 30 は、このアクセント値を、前後の音素や前後の音節との関係に基づくルール等によって修正する。

【0043】

ハードディスク 44 の解析用辞書 50 中の音節の継続時間長の辞書には、図 9 に示すように、全ての音節とその継続時間長との関係が記述されている。CPU 30 は、この継続時間長の辞書を参照して、各音節の継続時間を取得する。さらに、この継続時間を、前後の音素や前後の音節との関係に基づくルール等によって修正する（ステップ S 4）。以上に基づいて、図 10 に示すような、音節ごとのテーブルを生成する。

【0044】

解析辞書 50 中の子音／母音の有声／無声辞書には、図 12 に示すように、全ての音素とその有声／無声の区別が記録されている。なお、図において、音素に付されたインデックスは、「V」が母音（有声音）、「CU」が子音の無声音、「CV」が子音の有声音を示している。CPU 30 は、この辞書を参照して、各音素「k」「o」「N」「n」「i」「c」「h」「i」「w」「a」のそれぞれについて

、有聲無聲の区別を行う。さらに、有聲音が無聲化するルールに基づいて、無聲化する部分を決定する。このようにして、各音素につき、有聲無聲の判断を行う（ステップ S 5）。

【0 0 4 5】

次に、CPU 3 0 は、図 1 0 のテーブル（特にアクセント値）に基づいて、図 1 1 に示すような、基本周波数 F_0 の時間的変化を生成する（ステップ S 6）。なお、基本周波数が示されていない部分は、無聲音の部分である。

【0 0 4 6】

このようにして、基本周波数 F_0 の時間的変化を決定すると、次に、有聲音源強度 A_v 、無聲音源強度 A_f の時間的変化を決定する（ステップ S 7）。解析辞書 5 0 の音源強度辞書には、図 1 3 に示すように、全ての音節に対応づけて音源強度の時間的変化が記録されている。CPU 3 0 は、これを参照して「ko」「N」「ni」「chi」「wa」の各音節について、有聲音源強度 A_v 、無聲音源強度 A_f を得る。さらに、アクセント値や前後の音の環境等に基づいて、取得した音源強度を修正する。また、ステップ S 4 で決定した継続時間長に合致するように、音源強度の時間的変化を修正する。

【0 0 4 7】

CPU 3 0 は、各音節について、音声データベース 4 8 から、サンプル音声波形を取得する。音声データベース 4 8 には、図 1 4 に示すように、実際に発話されたサンプル音声波形が、音節ごとに区切られて音韻情報が付与されて記録されている。また、それぞれの音節について、音源強度の時間的変化、基本周波数の時間的変化、継続時間長、ピッチマーク、ゼロクロスマークが、併せて記録されている。ここで、ピッチマークとは、各ピッチにおけるピーク値（図 1 の M 参照）の位置に付されたマークである。また、ゼロクロスマークとは、各ピッチのマイナスピーク（図 1 の α 参照）の直前のゼロクロス点に付されたマークである。この実施形態では、ピッチマーク、ゼロクロスマークを時刻によって示している。

【0 0 4 8】

なお、多くのサンプル波形が記録されているので、同じ音節「ko」に対しても

、複数のサンプル波形が存在することとなる。CPU 30は、前後のつながり等から、各音節に対してもっとも好ましいサンプル波形を取得する（ステップS8）。

【0049】

次に、CPU 30は、各音節のそれぞれについて、音声データベース48から取得したサンプル音声波形の継続時間長が、ステップS4において決定した継続時間長に等しくなるように、サンプル音声波形を修正する（ステップS9）。これは、数ピッチの波形を重複させたり（同じ波形を挿入したり）、削除したりすることによって行う。

【0050】

つづいて、CPU 30は、各音節のそれぞれについて、音声データベース48から取得したサンプル音声波形の音源強度の時間的変化が、ステップS7において決定した強度の時間的変化に等しくなるように、サンプル音声波形を修正する（ステップS10）。

【0051】

さらに、CPU 30は、各音節のそれぞれについて、音声データベース48から取得したサンプル音声波形の基本周波数 F_0 の時間的変化が、ステップS6において決定した基本周波数 F_0 の時間的変化に等しくなるように、サンプル音声波形を修正する（ステップS11）。

【0052】

図7に、ピッチ変更処理プログラムのフローチャートを示す。なお、ピッチ変更処理は、有声音の波形についてのみ行う。無声音の波形には、規則的な繰り返し波形がないためである。

【0053】

まず、CPU 30は、サンプル音声波形における（対象とする音節の）最初のピッチの基本周波数を、音声データベース48の基本周波数の変化のデータから取得する。次に、ステップS6において決定した基本周波数の変化に基づいて、対応する部分の基本周波数を取得する。CPU 30は、両者が合致しているか否かを判断する（ステップS22）。合致していれば、当該ピッチについてピッチ

変更をおこなう必要がないので、ステップ S 2 6 に進む。

【0 0 5 4】

合致していなければ、サンプル音声波形のピッチを長くすべきか（基本周波数を下げるべきか）、短くすべきか（基本周波数を上げるべきか）を判断する（ステップ S 2 3）。その判断結果に基づいて、ピッチを長くするか（ステップ S 2 5）、または短くするか（ステップ S 2 4）の処理を行う。

【0 0 5 5】

CPU 3 0 は、対象となっているピッチにおけるマイナスピーク直前のゼロクロスを見いだす。ゼロクロス点は、図 1 4 に示すように音声データベースに記録されているので、容易に得ることができる。

【0 0 5 6】

ピッチを長くする場合には、図 1 5 に示すように、このゼロクロス点にゼロ値を挿入する。

【0 0 5 7】

また、ピッチを短くする場合には、ゼロクロスの前後にほぼゼロ値に近い値の部分があれば、当該部分を必要なだけ削除する。ゼロクロスの前後にほぼゼロ値に近い値の部分がないければ、図 1 6 に示すような操作を行って、ピッチを短くする。まず、マイナスピークの前、 $2N-1 \sim N$ のフレームに、 $2N-1$ 前で窓値 1、 N 前で窓値 0 となるハニング窓をかける。同様に、マイナスピークの前 $N-1 \sim$ マイナスピークのフレームに、マイナスピークで窓値 1、マイナスピークの前 $N-1$ で窓値 0 となるハニング窓をかける。これら 2 つの窓処理後の成分を加算したものを、修正後の波形として用いる。これにより、 $2N$ フレームを N フレームに短くすることができる。

【0 0 5 8】

なお、上記窓処理において、ゼロクロス近傍に窓値 0 の部分がくるようにし、ゼロクロスより離れるほど窓値が 1 に近づくようにしてもよい。これにより、ゼロクロスより離れた地点では、窓値を「1」、つまり波形をそのまま保存するようにし、ゼロクロスでは、窓値を「0」、つまり実質的に波形を削除することができる。つまり、影響の少ないと思われるゼロクロス付近での加工値

を大きくし、自然性を保持することができる。

【 0 0 5 9 】

上記のようにしてピッチ変更処理を行うと、CPU 30は、全てのピッチについて以上の処理を行ったかどうかを判断する（ステップS 2 6）。当該音節について未処理のピッチがあれば、次のピッチを対象として（ステップS 2 7）、ステップS 2 2以下の処理を繰り返す。全てのピッチについて処理が終了すれば、当該音節についての処理を終了する。なお、ピッチ変更によって継続時間長に微調整が必要となれば、これも行う。また、ピッチ変更処理は、サンプル波形として選択された全ての音節に対して行われる。

【 0 0 6 0 】

上記のようにしてピッチ変更処理が終了すると、図6のステップS 1 2を実行する。ステップS 1 2においては、音節ごとにサンプル波形を修正して得られた波形を接続し、合成音声波形を得る。最後に、CPU 30は、この音声合成波形を、サウンドカード54に出力する。サウンドカード54は、これをアナログ信号に変換し、スピーカ56から音声として出力する。

【 0 0 6 1 】

(4) 音声データベースに関する他の実施形態

上記実施形態においては、多くのサンプル波形を音節を音声単位として記録した音声データベース（音声コーパス）を用いている。しかしながら、音素を音生単位として記録したものを用いてもよい。また、音節に加えて、複数の音節が明瞭な区分に乏しく連続している場合にはこれら音節を1かたまりとしてあつかうようにしてもよい（拡張音節）。その定義表を図17に示す。軽音節よりも重音節が優先して切り出され、重音節よりも超重音節が優先して切り出される。つまり、超重音節に該当する場合には、その一部が重音節として切り離されることはない。同様に、重音節に該当する場合には、その一部が音節として切り離されることはない。このように、「複数の音節が明瞭な区分に乏しく連続している場合にはこれら音節を1かたまりとしてあつかう」ことにより、接続による不自然さを排除することができる。なお、少なくとも、軽音節と重音節を用いることが好ましい。

【0062】

また、上記実施形態では、音声コーパスを用いているが、各音節（または音素、拡張音節）につき1つの音声波形データを記録した音声データベースを用いてもよい。さらには、各音節（または音素、拡張音節）ごとに、1つのピッチだけの音声波形データを記録した音声データベースを用いてもよい。

【0063】

また、上記実施形態では、音声データベースにゼロクロスマークを記録している。しかしながら、ゼロクロスマークを記録せず、ピッチマーク等に基づいて、処理の都度、探し出すようにしてもよい。

【0064】

(5)ピッチ変更処理に関する他の実施形態

上記実施形態では、ゼロクロスにおいて、実質的なゼロ値を挿入、削除することにより、ピッチ変更を行っている。しかしながら、マイナスピークに向かう減少部分（図1のγ参照）において、時間的圧縮、時間的伸長を行って、ピッチ変更をしてもよい。一般的には、時間的圧縮や、時間的伸長により、ピッチ変更に関係しない余分な周波数成分がもたらされる可能性がある。しかし、このマイナスピークに向かう減少部分は、単調減少であり、多くの周波数成分を含んでいないので、加工による音声品質への影響が少ないと考えられる。

【0065】

なお、上記において、マイナスピークに向かう減少部分において、全体的に一樣に時間的圧縮、時間的伸長を行うのではなく、ゼロクロス付近において集中的に時間的圧縮、時間的伸長を行い、ゼロクロスから離れるにしたがって、その程度を小さくするようにしてもよい。

【0066】

2. 第2の実施形態

図18に、この発明の第2の実施形態による音声合成装置の全体構成を示す。この実施形態では、音声波形合成手段16は、文字列解析手段2、音声単位波形生成手段90、波形結合手段22を備えている。音声データベース18には、音

声単位（たとえば音節）を生成するための 1 ピッチ分の音声波形が、ピッチを少しずつ変えて記録されている。たとえば、「あ」という音節を生成するために必要な 1 ピッチ分の波形が、1ms 程度、ピッチ長を順次違えて多数記録されている。他の全ての音節（有声音）についても同様に記録されている。なお、無声音のためには、雑音波形が記録されている。

【0067】

入力された文字列は、解析用辞書 20 を参照して、文字列解析手段 2 によって形態素解析がなされ、音声単位に分離される。さらに、解析用辞書 20 を参照し、前後の音の環境等を考慮して、各音声単位ごとに、無声音／有声音の区別、継続時間長、強度の時間的変化、基本周波数の時間的変化を算出する。

【0068】

音声単位波形生成手段 90 は、各音声単位を生成するために必要な 1 ピッチ分の波形を音声データベースから取得する。この際、文字列解析手段 2 からの基本周波数の時間的変化に基づいて、各時間ごとに適切なピッチ長を持つ波形を選択して取得する。さらに、音声単位波形生成手段 90 は、これら複数個のピッチの波形を、継続時間長、強度の時間的変化を参照して、変形して接続し、音声単位の波形を生成する。なお、無声音については、雑音波形に基づいて、波形を生成する。

【0069】

上記のようにして生成された各音声単位の音声波形は、波形結合手段 22 によって接続され、音声波形データとして出力される。

【0070】

アナログ変換手段 14 は、この音声波形データをアナログ変換して音声信号として出力する。

【0071】

図 18 の音声合成装置を、CPU を用いて実現した場合のハードウェア構成例は、図 4 と同様である。この実施形態では、音声単位（音節など）の波形を、1 ピッチ分の波形を接続して合成するようにしている。また、そのため、音声データベース 18 には、図 19 に示すように、各音声単位について、当該音声単位を

生成するために必要な 1 ピッチ分の音声波形データを、種々のピッチで用意している。また、ピッチ長を異ならせるために、マイナスピーク直前のゼロクロス点において、ゼロ値が挿入されたものとなっている。

【0 0 7 2】

なお、この実施形態においても、第 1 の実施形態のようにして、処理の都度、ピッチを変更する処理を行うようにしてもよい。このようにすれば、音声データベースには、1 種類のピッチ長のデータを記録しておくだけでよい。

【0 0 7 3】

また、第 1 の実施形態において述べた他の形態は、この第 2 の実施形態においても適用することができる。

【0 0 7 4】

3. その他の実施形態

上記実施形態では、文字列解析手段 2 の解析結果に従って、ピッチの変更を行うようにしている。しかしながら、操作者の入力した指令にしたがってピッチ変更を行うようにしてもよい。

【0 0 7 5】

図 2 0 に、このような指令を入力するための画面表示例を示す。図 2 1 に、ハードディスク 4 4 に記録された指令入力判断のためのプログラムをフローチャートにて示す。

【0 0 7 6】

操作アイコン 2 0 0（上向きの矢印）を、マウス 3 4 によってクリックすると、音声の強度およびピッチ周波数が全体的に上がるように修正が行われる（ステップ S 5 0、S 5 3）。同様に、操作アイコン 2 0 4（下向きの矢印）をクリックすると、音声の強度およびピッチ周波数が全体的に下がるように修正が行われる（ステップ S 5 0、S 5 2）。また、操作アイコン 2 0 6（左向きの矢印）をクリックすると、例えば、数個のピッチを削除するなどして発話音声長が短くなるように修正が行われ（ステップ S 5 0、S 5 1）、操作アイコン 2 0 2（右向きの矢印）をクリックすると、例えば、数個のピッチを重複させるなどして発話

音声長が長くなるように修正が行われる（ステップ S 5 0、S 5 4）。

【0 0 7 7】

なお、ピッチの修正は、上記の実施形態に示した方法によって行うことが好ましいが、他の方法であってもよい。

【0 0 7 8】

このように、一对の矢印（上方向と下方向または左方向と右方向）に、それぞれ、変更方向が逆の処理を対応づけるようにしている。したがって、処理内容が、直感的に理解しやすく、音声に対する処理指令の入力が容易となる。

【0 0 7 9】

なお、上記実施形態では、画面上のアイコンを用いているが、これに代えて、矢印形状を表した（外形が矢印形状または矢印形状が表示された）入力スイッチ等を用いてもよい。

【0 0 8 0】

また、上記実施形態では、上向きの矢印、下向きの矢印のそれぞれに、音声強度、ピッチ周波数の 2 つの処理を対応づけているが、音声強度、ピッチ周波数、発話継続時間のいずれか 1 つ、または、いずれか 2 つ、さらには、3 つを対応づけるようにしてもよい。左向き矢印、右向き矢印についても同様である。加えて、斜め向きの矢印を設け、これをクリックすることにより、縦方向の矢印と横方向に矢印に対応づけられた処理の双方を実行するようにしてもよい。

【0 0 8 1】

4. その他

上記実施形態では、図 3、図 1 8 の各機能を CPU を用いて実現した場合について説明したが、その一部又は全部をハードウェアロジックによって構成してもよい。

【図面の簡単な説明】

【図 1】

「あ」の音声波形の一部を示す図である。

【図 2】

多数の「あ」の音声波形を重ねて示す図である。

【図 3】

本発明の一実施形態による音声合成装置の全体構成を示す図である。

【図 4】

図 3 の装置のハードウェア構成を示す図である。

【図 5】

音声合成処理プログラムのフローチャートである。

【図 6】

音声合成処理プログラムのフローチャートである。

【図 7】

ピッチ変更処理プログラムのフローチャートである。

【図 8】

単語辞書を示す図である。

【図 9】

音節の継続時間長の辞書を示す図である。

【図 1 0】

解析テーブルを示す図である。

【図 1 1】

決定した基本周波数の時間的变化を示す図である。

【図 1 2】

子音／母音の有声／無声辞書を示す図である。

【図 1 3】

音源強度辞書を示す図である。

【図 1 4】

音声データベースを示す図である。

【図 1 5】

ゼロ値の挿入によるピッチ変更を示す図である。

【図 1 6】

ゼロ値の削除によらないピッチ短縮を示す図である。

【図 1 7】

拡張音節の定義を示す図である。

【図 1 8】

第 2 の実施形態の全体構成を示す図である。

【図 1 9】

音声データベースの内容を示す図である。

【図 2 0】

操作のためのアイコンを示す図である。

【図 2 1】

指令入力判断のためのプログラムのフローチャートである。

【図 2 2】

音声のピッチを示す図である。

【図 2 3】

PSOLA法によるピッチ変更処理を説明するための図である。

【図 2 4】

PSOLA法による処理の影響を示すための図である（原波形）。

【図 2 5】

PSOLA法による処理の影響を示すための図である（ハニング窓左側）。

【図 2 6】

PSOLA法による処理の影響を示すための図である（ハニング窓右側）。

【図 2 7】

PSOLA法による処理の影響を示すための図である（合成波形）。

【図 2 8】

PSOLA法によって生じるエコーを説明するための図である。

【符号の説明】

1 4 . . . アナログ変換手段

1 6 . . . 音声合成手段

1 8 . . . 音声データベース

特平 1 1 - 2 8 5 1 2 5

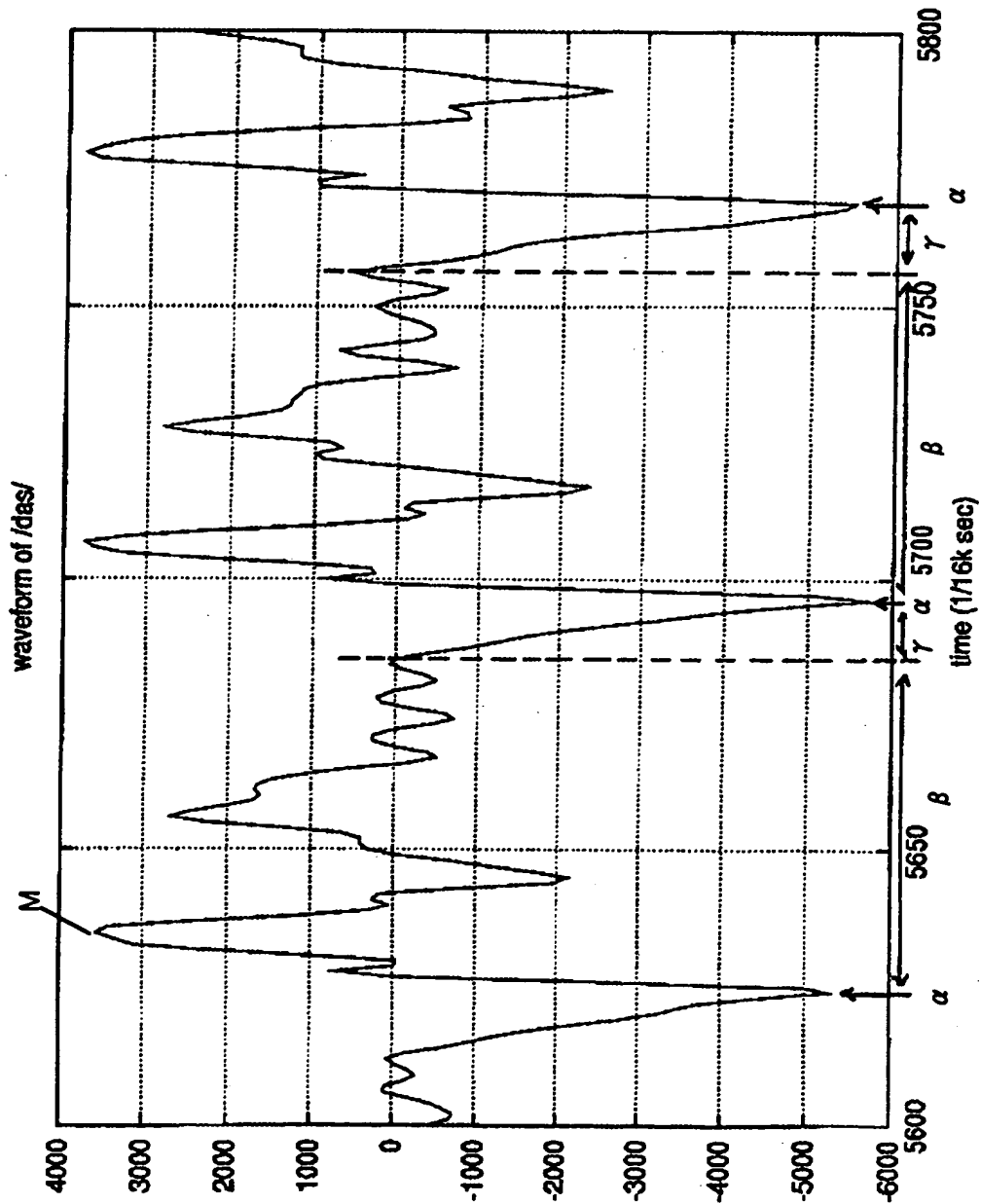
2 0 . . . 解析用辞書

【書類名】 図面

【図 1】

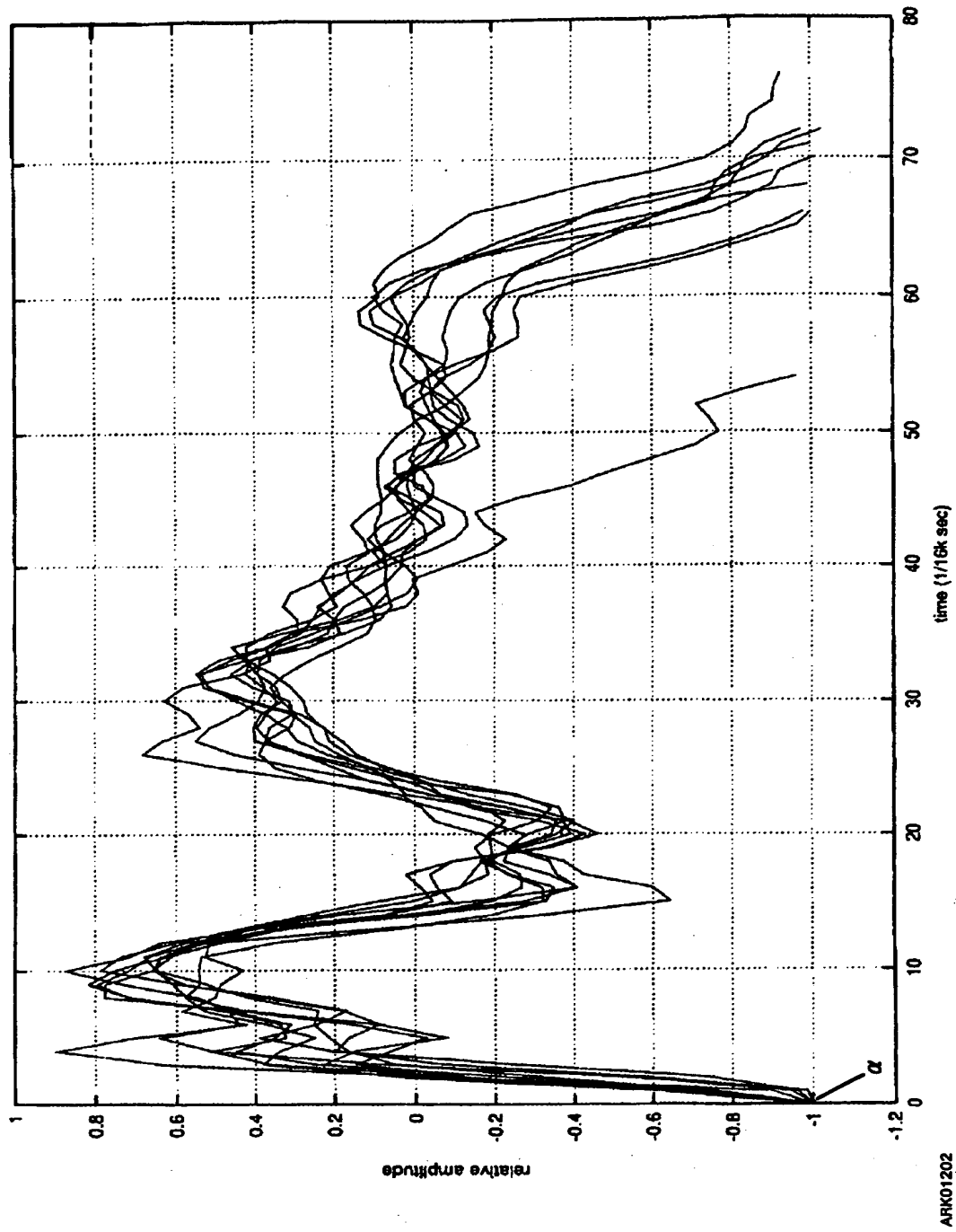
ARK012

図 1



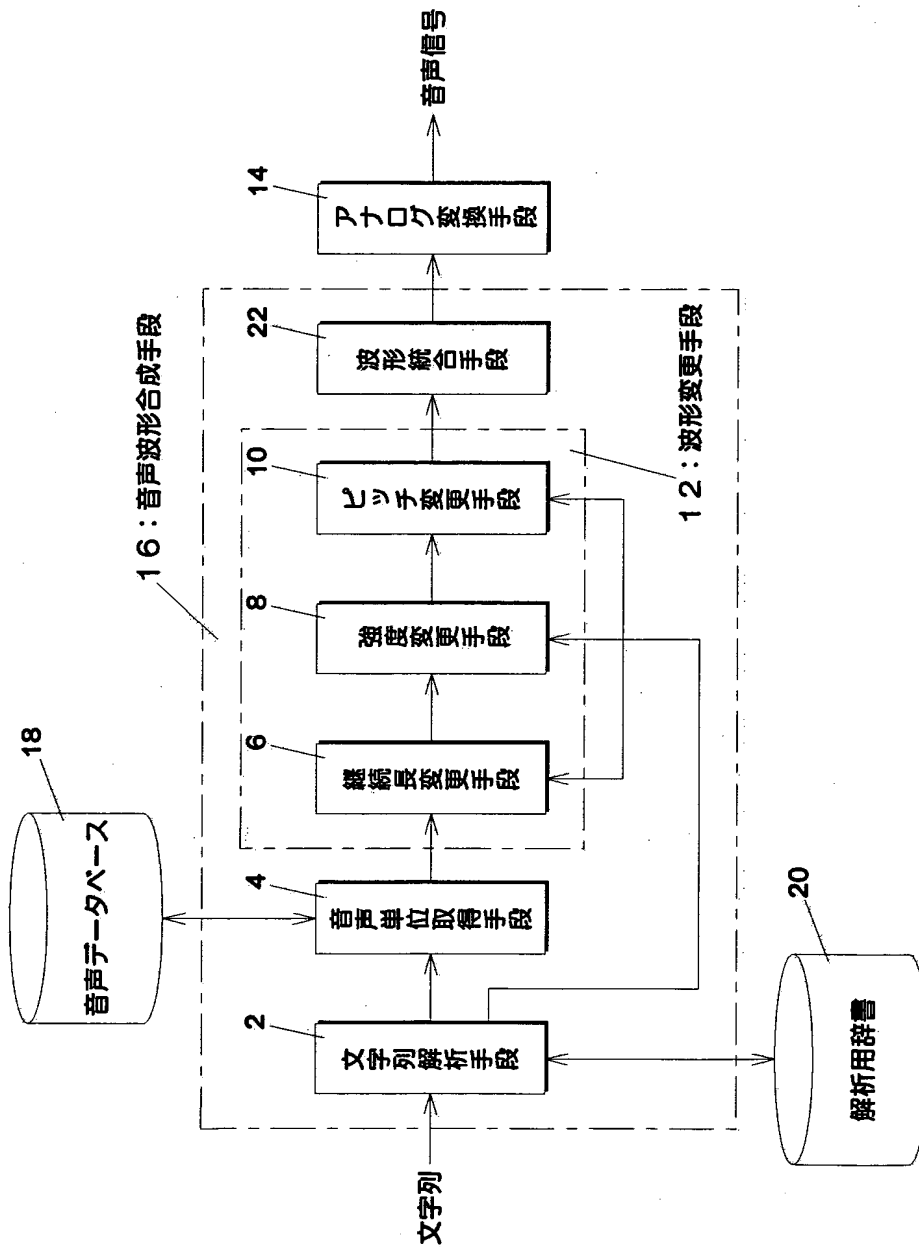
ARK01201

【図 2】



【図 3】

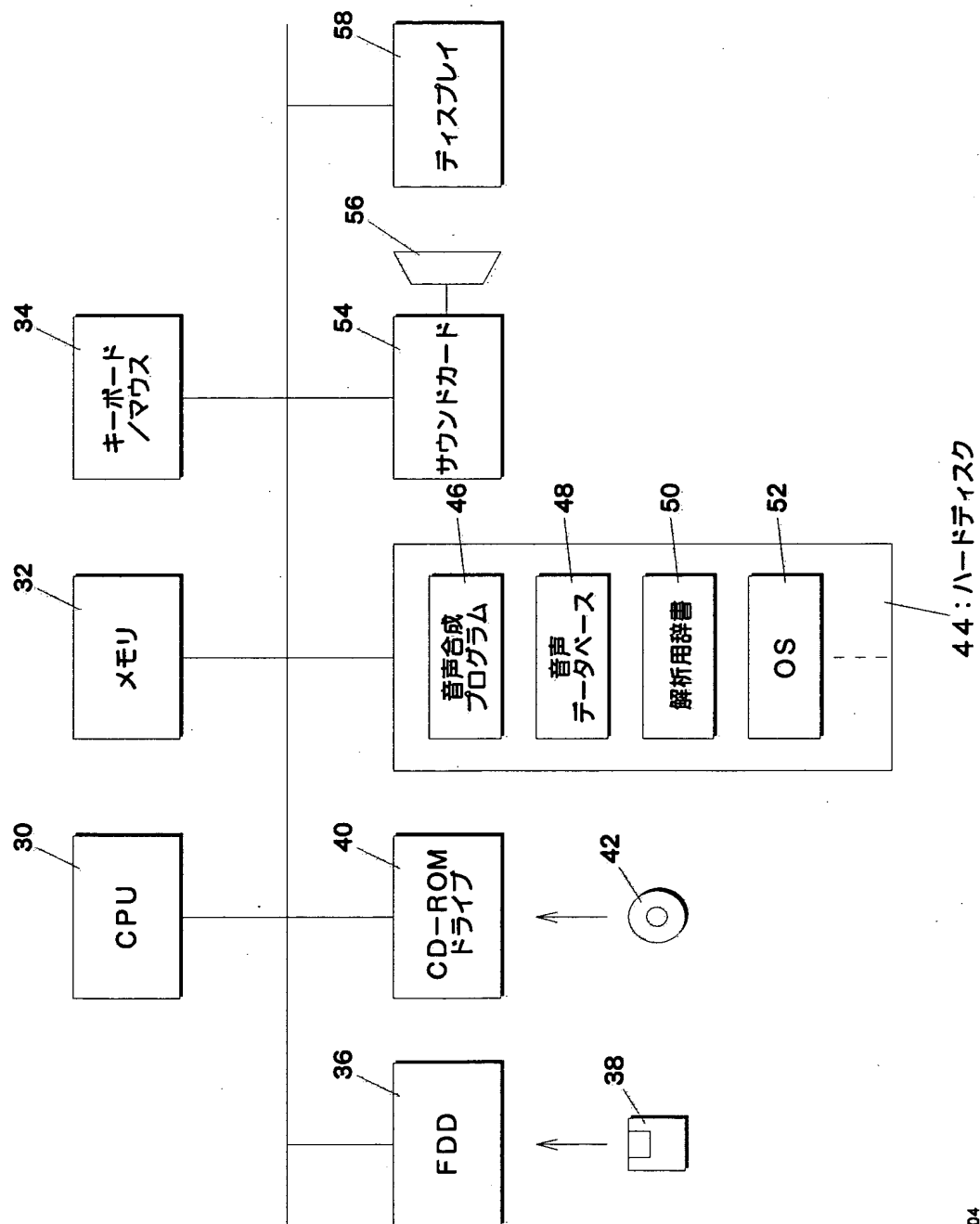
ARK012
図 3



ARK01203

【図 4】

ARK012
図 4

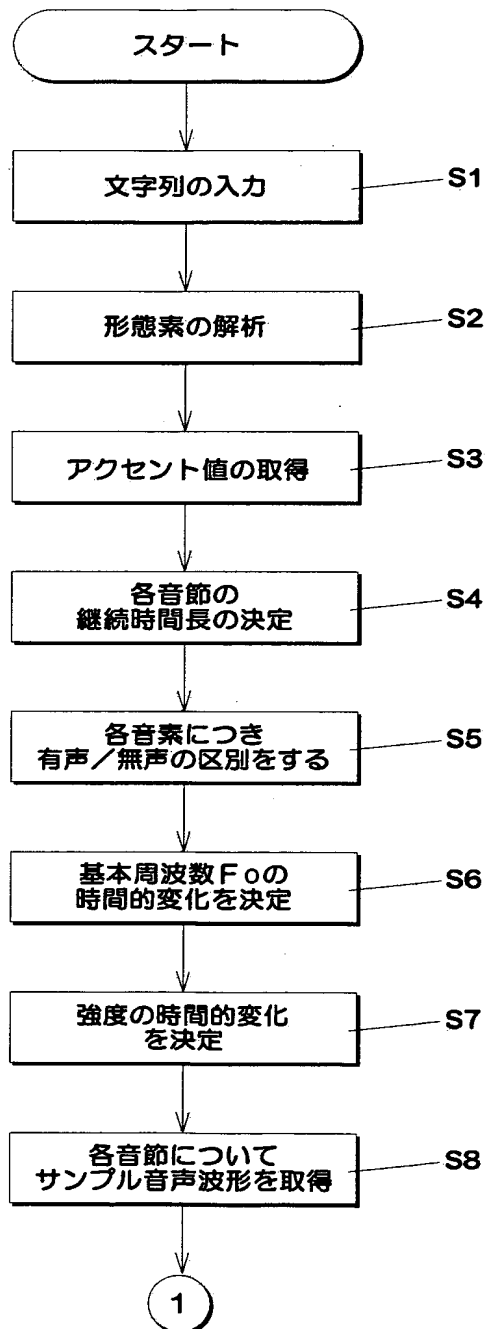


ARK01204

【図 5】

ARK012
図5

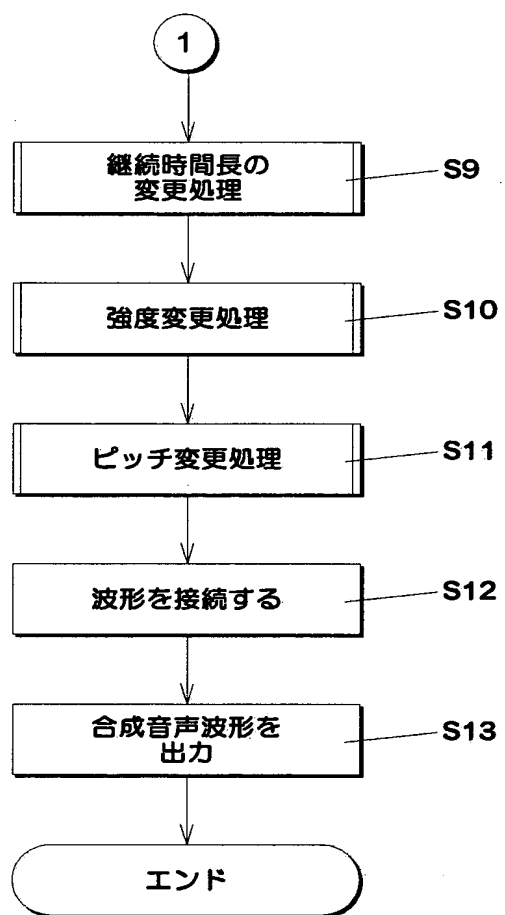
音声合成プログラム



ARK01205

【図 6】

ARK012
図 6

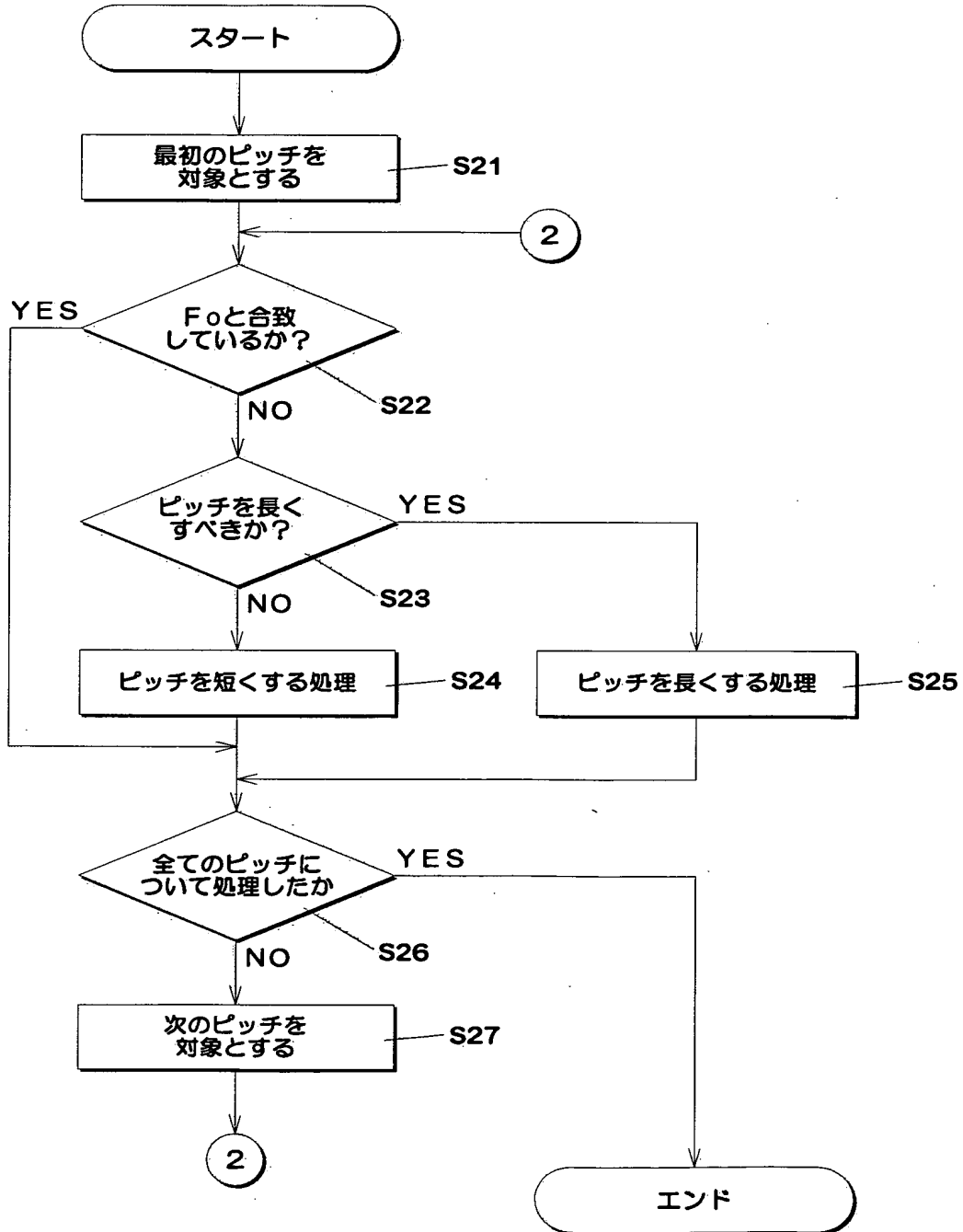


ARK01206

【図 7】

ARK012
図 7

ピッチ変更処理プログラム



ARK01207

【図 8】

ARK012
図8

単語辞書

単語	品詞	読み	----	音節	アクセント値
桜	名詞	s a k u r a	----	s a	1
				k u	5
				r a	0
さく	動詞	s a k u	----	s a	5
				k u	0
⋮	⋮	⋮		⋮	⋮
が	格助詞	g a	----	g a	2
⋮	⋮	⋮		⋮	⋮

ARK01208

【図 9】

ARK012
図 9

音節の継続時間長の辞書

音 節	継続時間長[m s]
a	1 1 0
i	1 1 4
u	9 0
l	l
k o	1 8 8
l	l

ARK01209

【図 1 0】

ARK012
図10

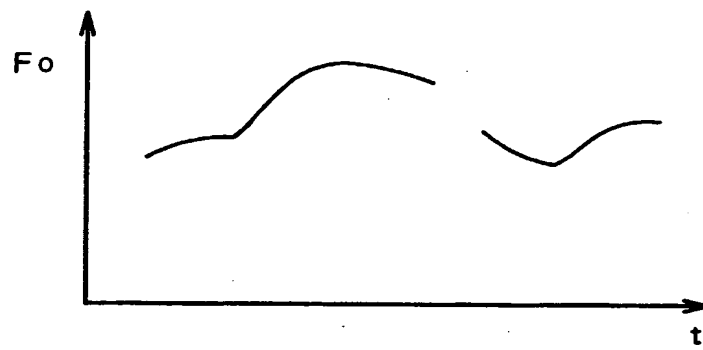
音節	音節連鎖	[ms] 時刻	[ms] 音節の継続時間長	アクセント値
ko	koN	0	188	0
N	Nni	188	92	5
ni	nich	280	212	5
chi	chiw	492	178	0
wa	wa	670	166	7

ARK01210

【図 1 1】

ARK012
図 1 1

ko Nn I chi wa



ARK01211

【図 1 2】

ARK012
図 1 2

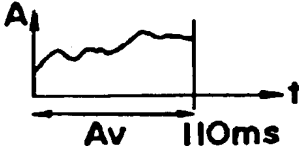
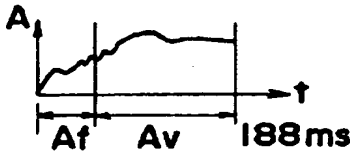
子音／母音の有声／無声辞書

音素	インデックス
a	V
i	V
⋮	⋮
k	CU
⋮	⋮
b	CV
⋮	⋮

ARK01212

【図 1 3】

音源強度辞書

音節	音源強度
a	
⋮	⋮
ko	
⋮	⋮

Av : 有声音源強度
Af : 無声音源強度

ARK01213

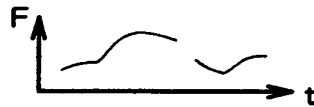
【図 1 4】



「こ」の
音源強度の変化



「こ」の
基本周波数の変化



「こ」の継続時間長

1 2 3ms

ピッチマーク

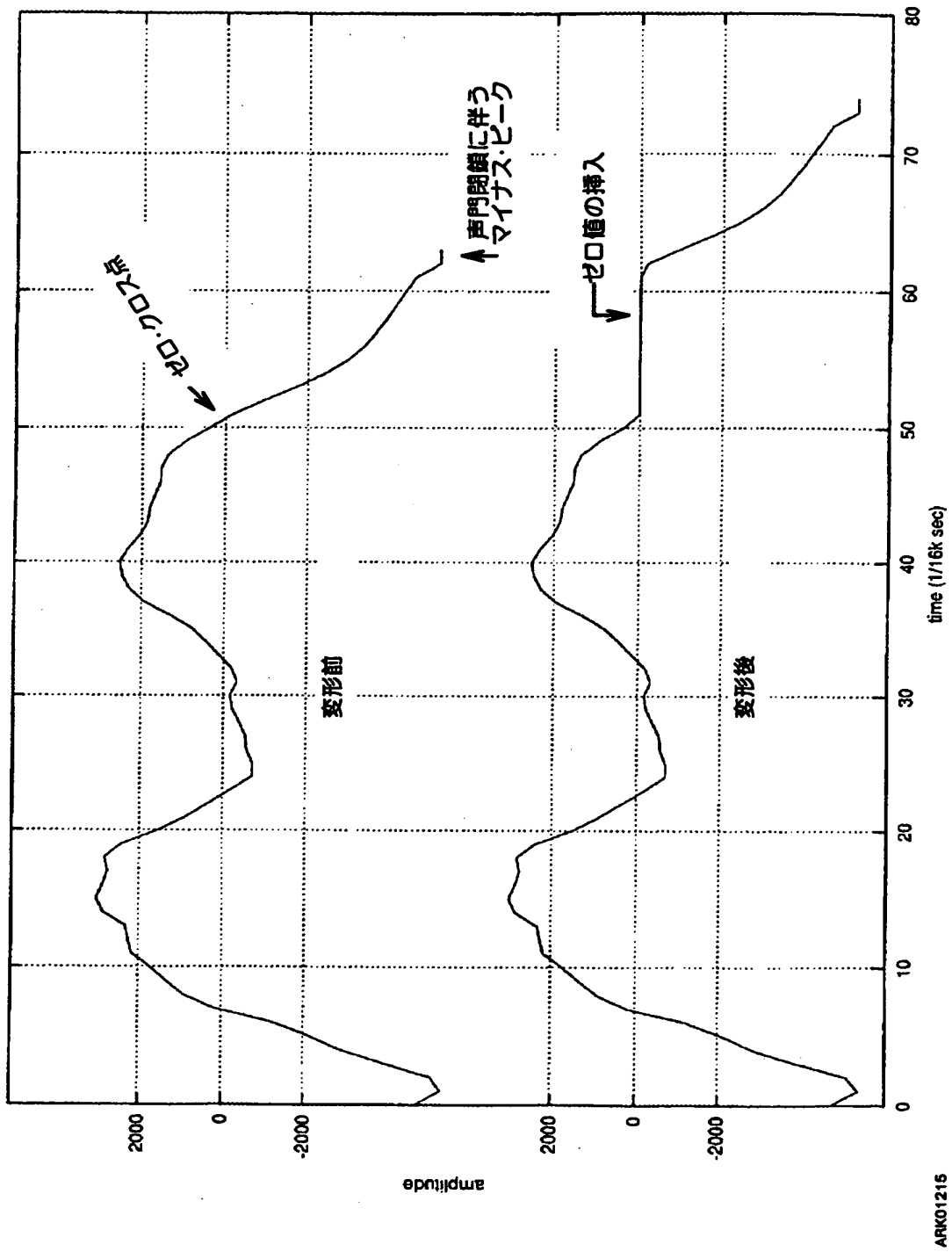
0、15、30、...

ゼロクロスマーク

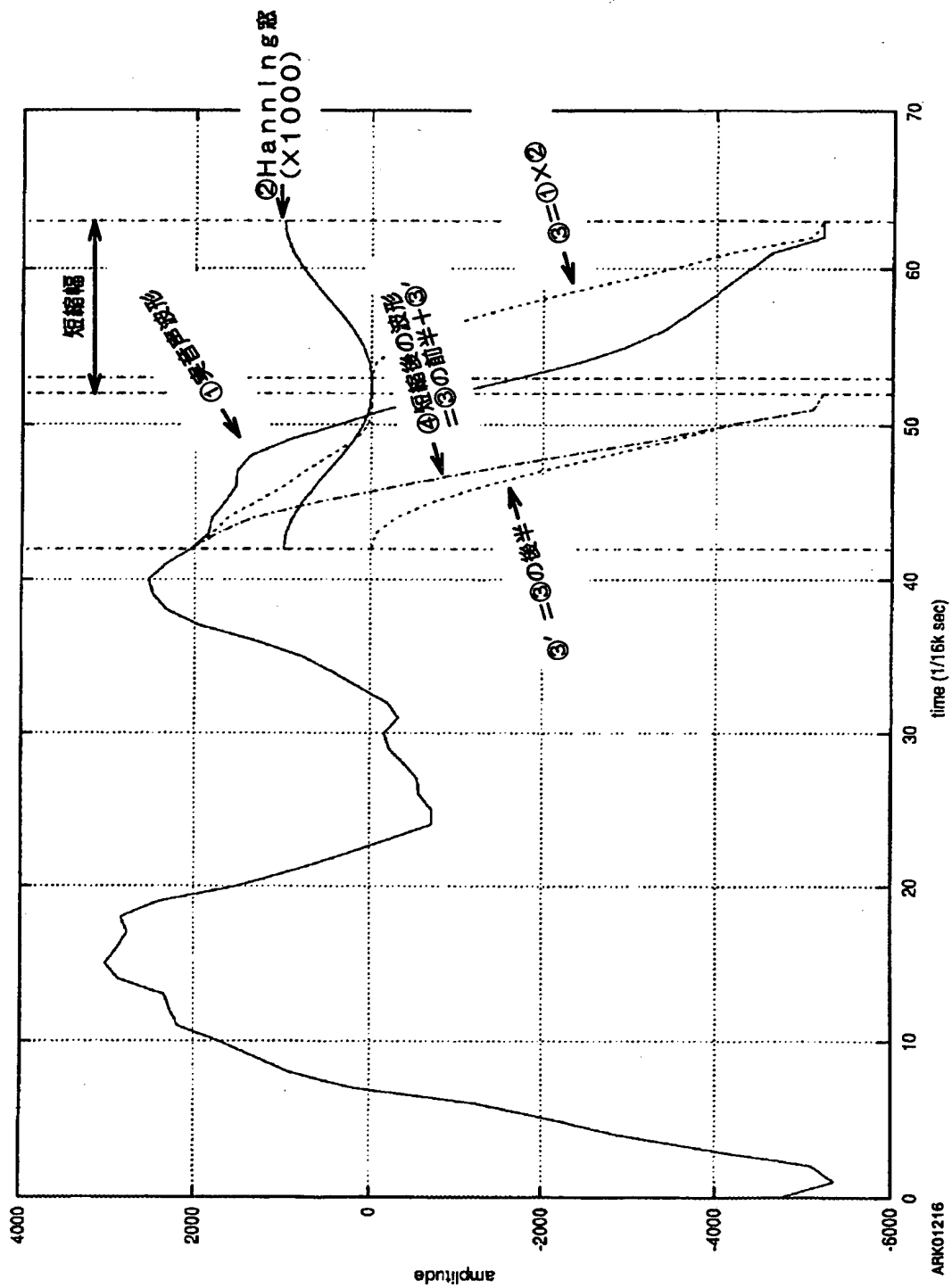
12、27、42、...

ARK01214

【図 1 5】



【図 1 6】



【図 17】

ARK012
図 17

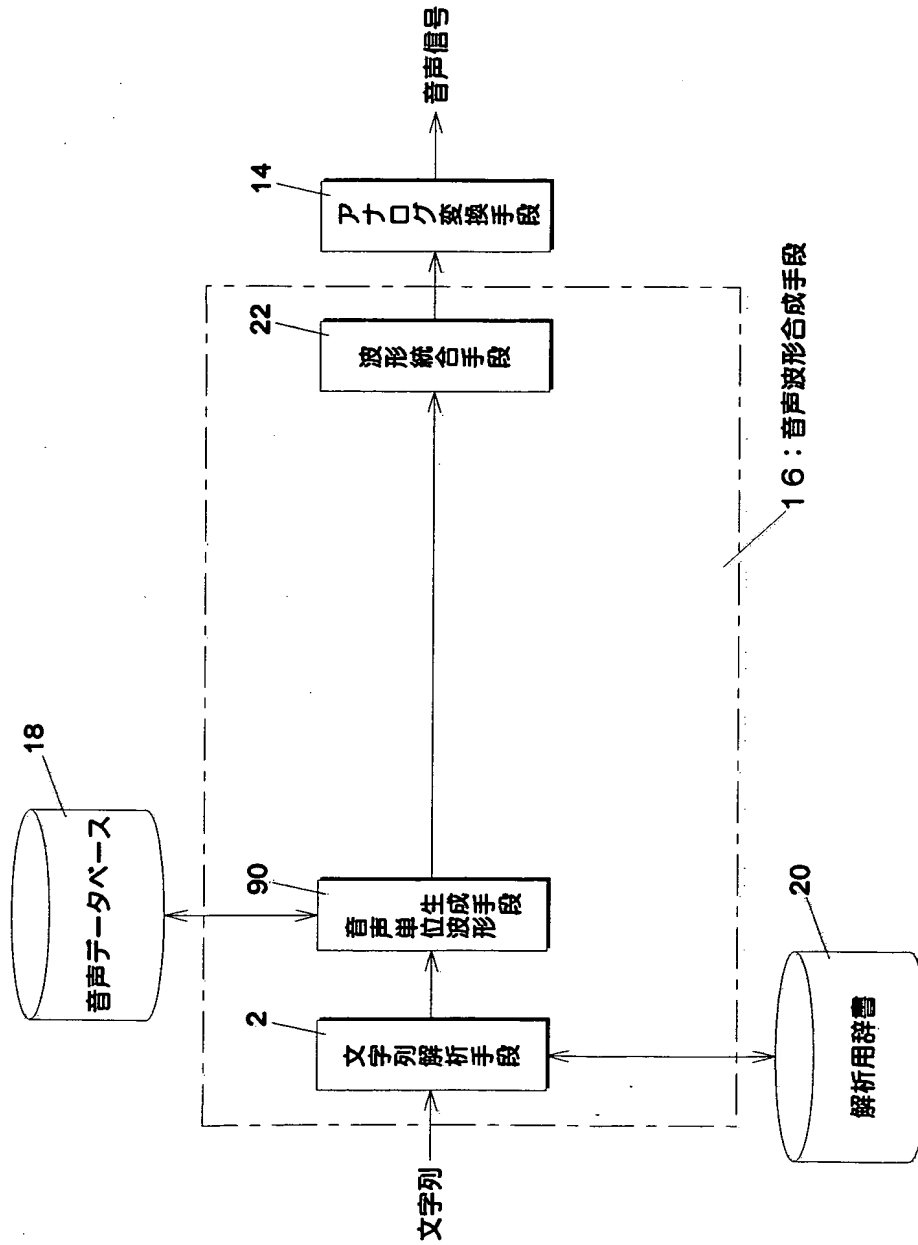
音節量	音節量タイプ	音節構造	例
1	軽音節	(C)(y)V	か, さ, た, な, は, ま, や, ら, … あ, い, う, え, お, ちえ, ぴゃ, …
2	重音節	(C)(y)VR (C)(y)VJ (C)(y)VN (C)(y)VQ	とー, やー, きゅー, ぴゅー, … かい, のう, あい, うい, ぴゅあ, … かん, あん, みゃん, ちゃん, … ちゅっ, りゃっ, じゃっ, ひゃっ, …
3以上	超重音節	(C)(y)VRN (C)(y)VRQ (C)(y)VJN (C)(y)VJQ (C)(y)VNQ など	ちえーん, じゅーん, あーん, … うーっ, ちえーっ, … さいん, ばうん, あうん, ちゃいん, … かいつ, だいつ, きゃいつ, ういつ, … どんっ, うんっ, ちゃんっ, …

C : 子音 (促音、拗音、撥音はふくまない)
y : 拗音
V : 母音 (長音、二重母音の第2要素を含まない)
R : 長音
J : 二重母音の第2要素
Q : 促音
N : 撥音
(X) : Xを含まなくとも、1以上含んでもよいことを示す

ARK01217

【図 1 8】

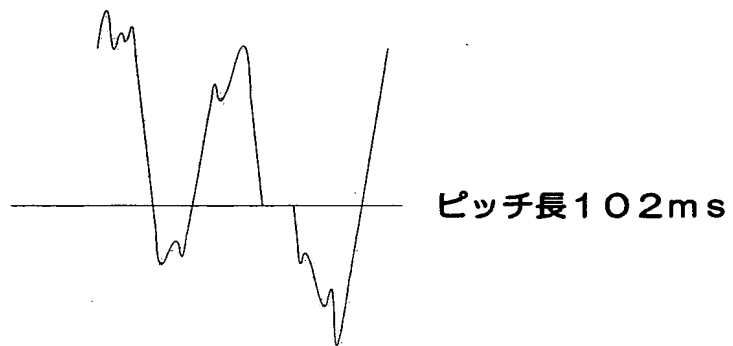
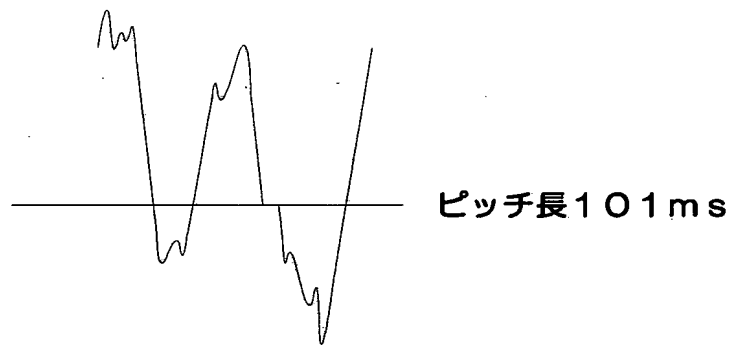
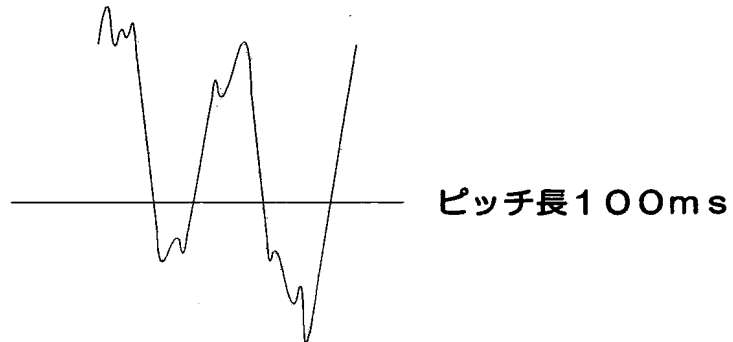
ARK012
図 18



ARK01218

【図 19】

ARK012
図 19

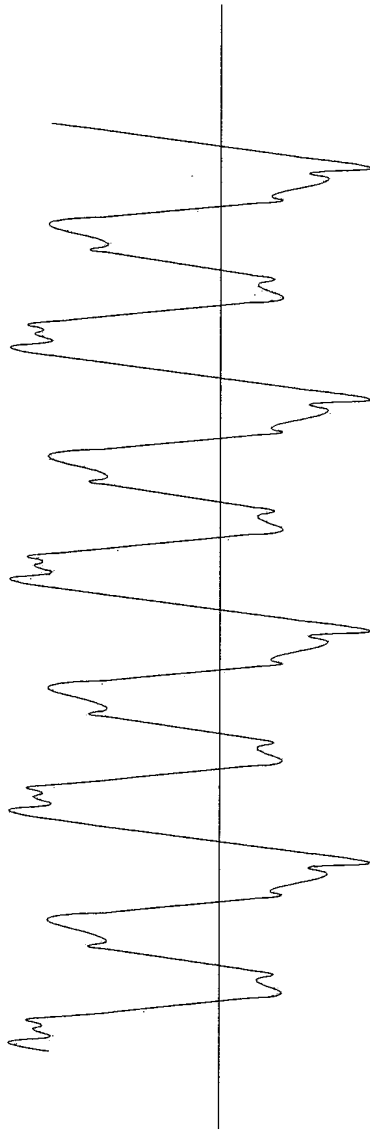


ARK01219

【図 2 0】

ARK012
図20

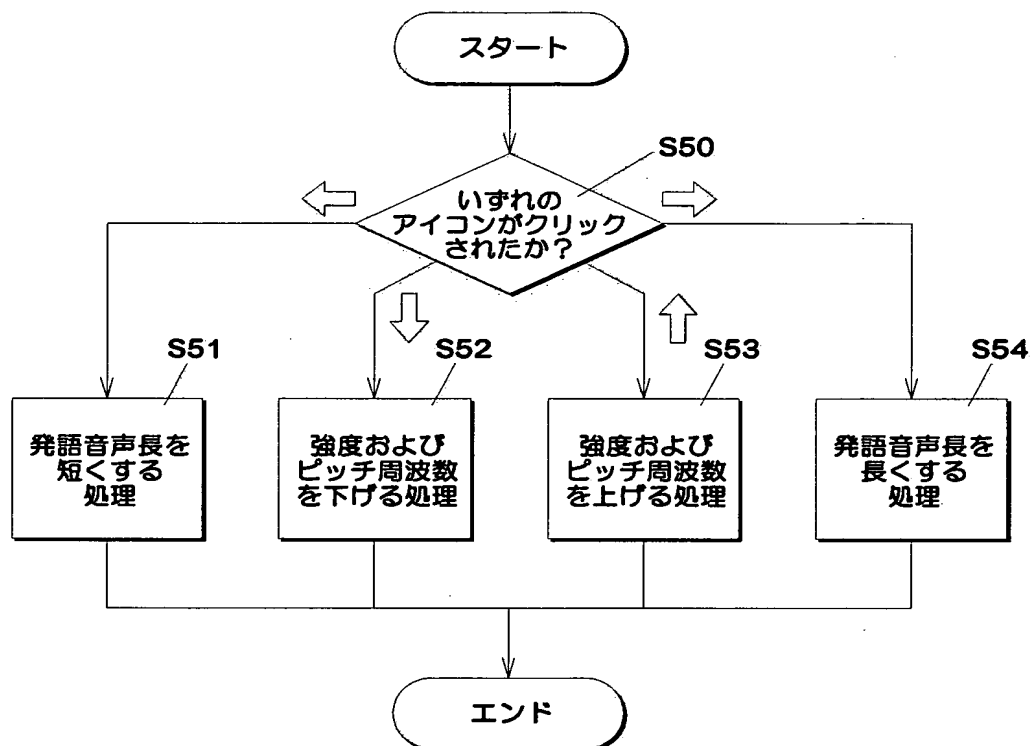
200
202
204
206



ARK01220

【図 2 1】

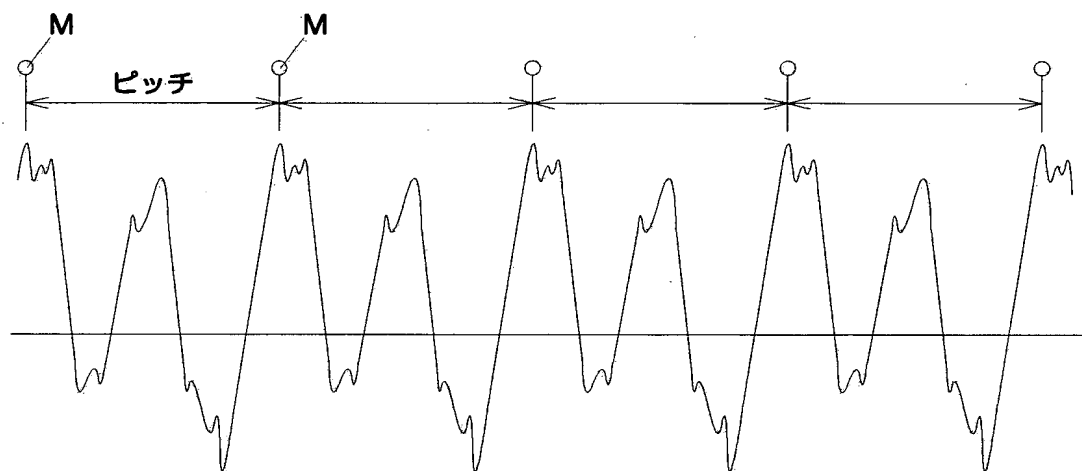
ARK012
図 21



ARK01221

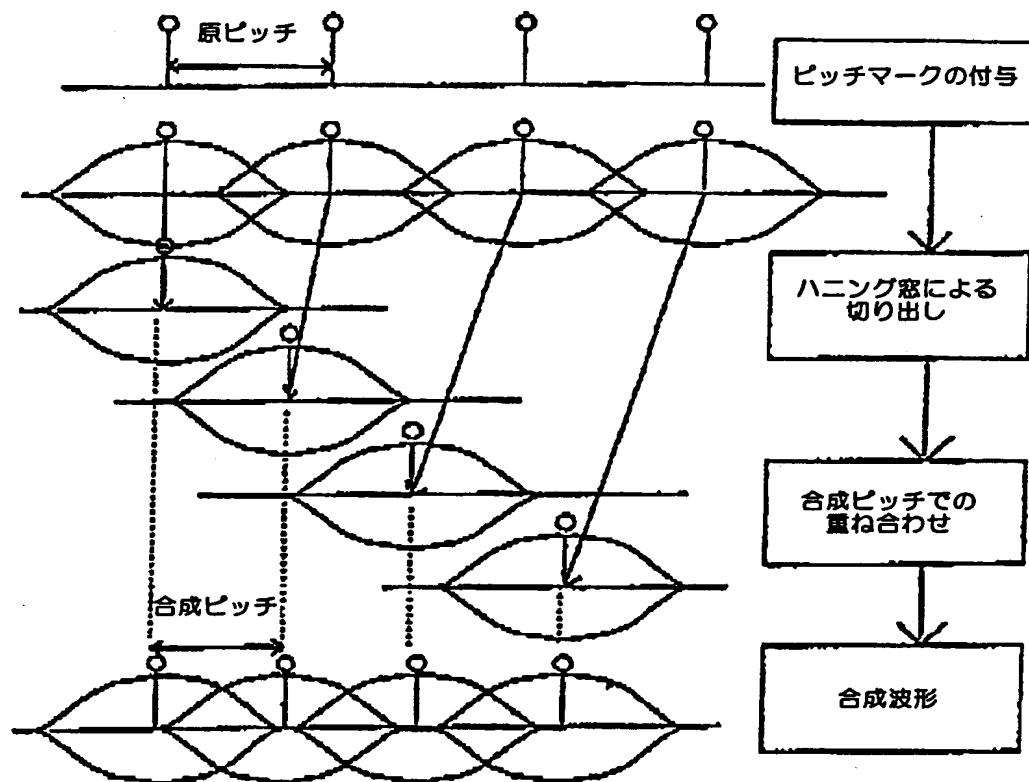
【図 2 2】

ARK012
図 22



ARK01222

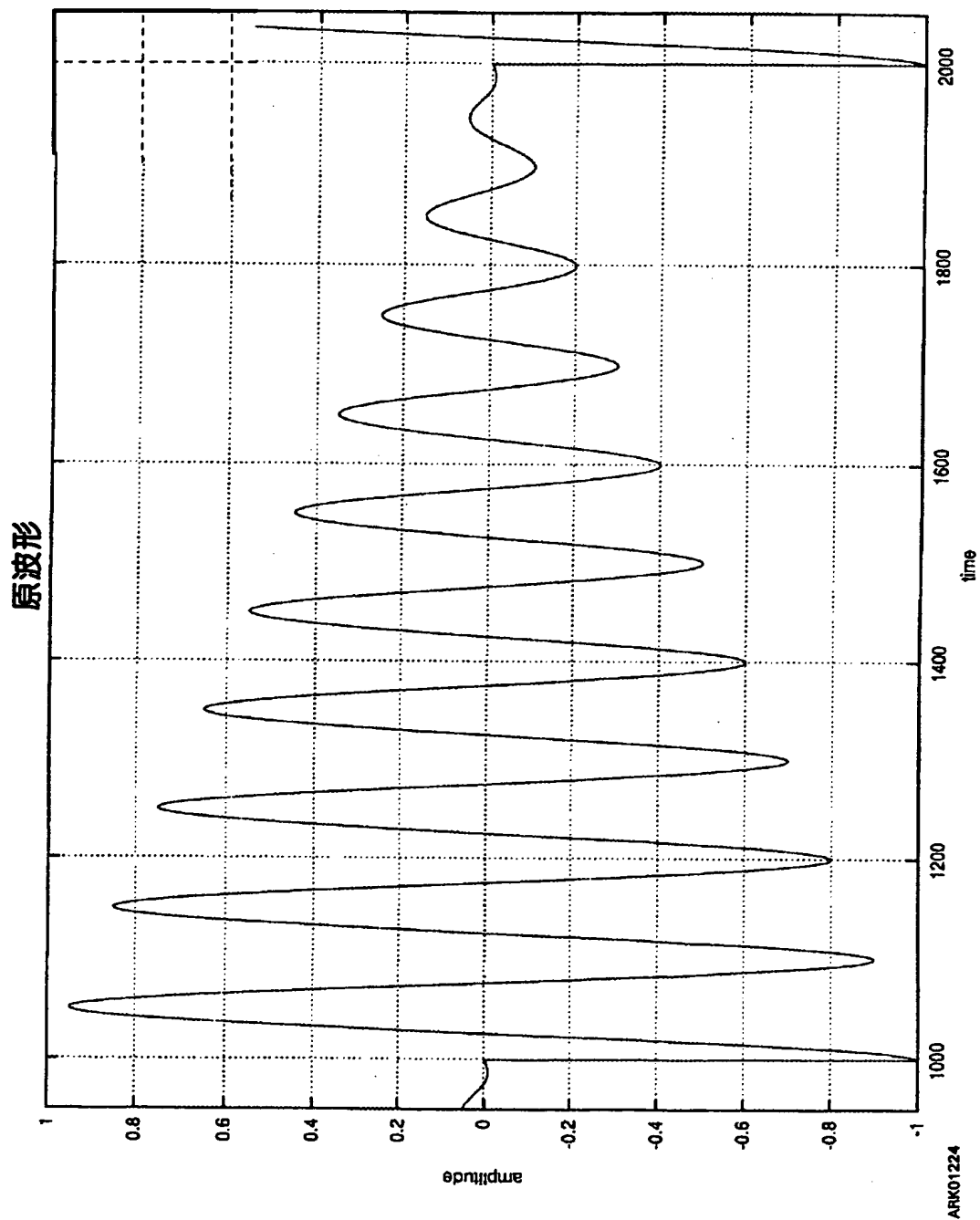
【図 2 3】



ARK01223

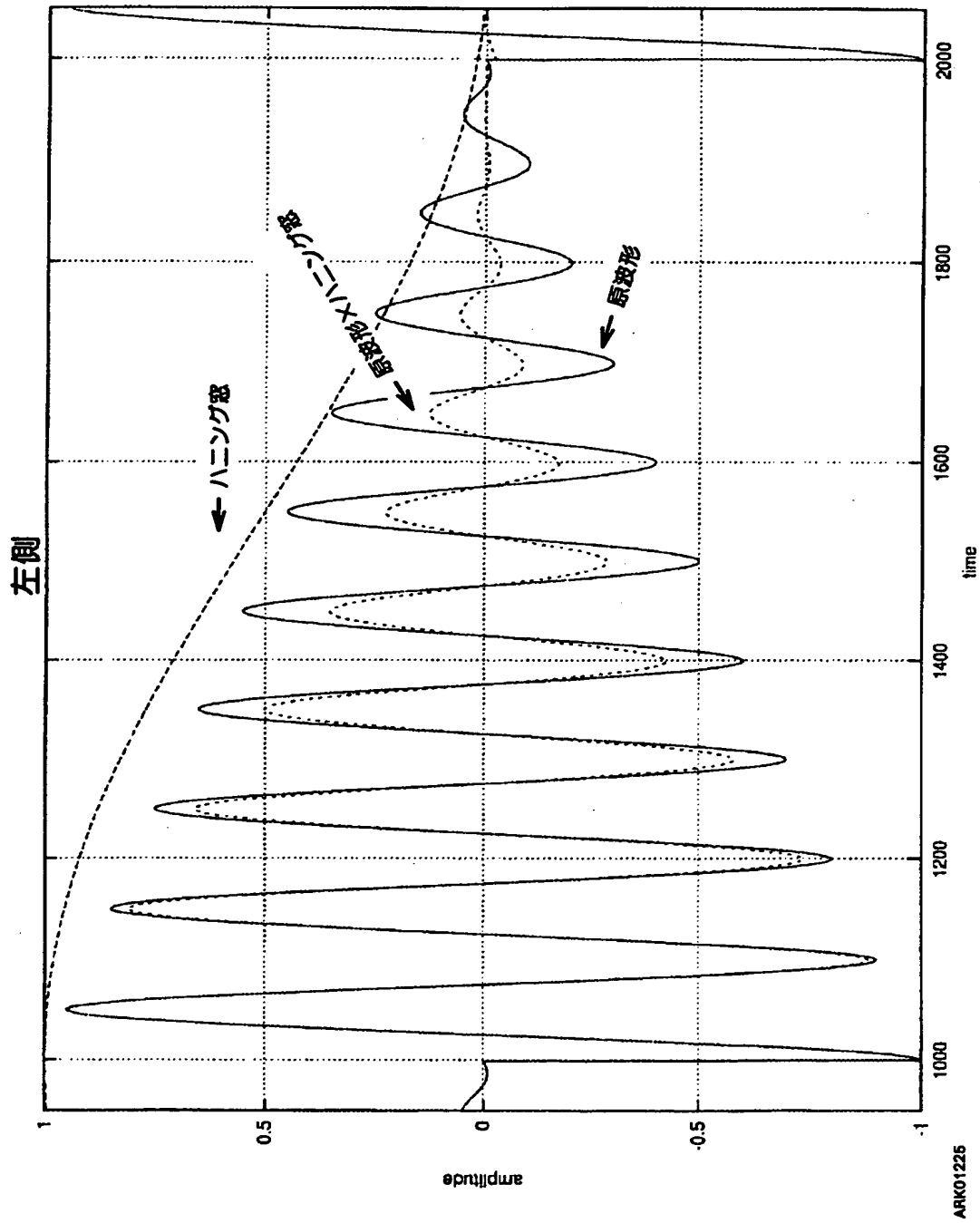
【図 2 4】

ARK012
図 24



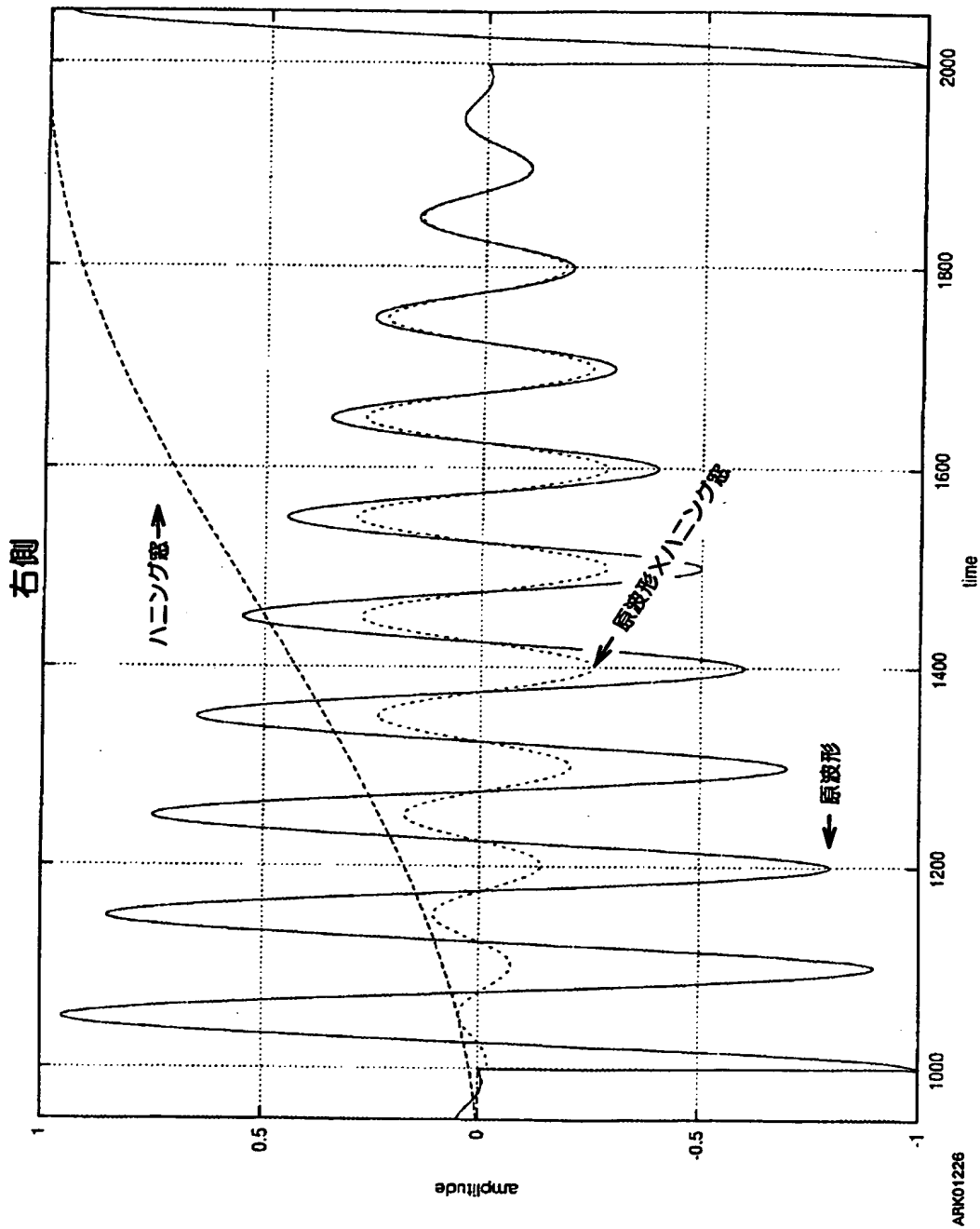
【図 2 5】

ARK012
図25



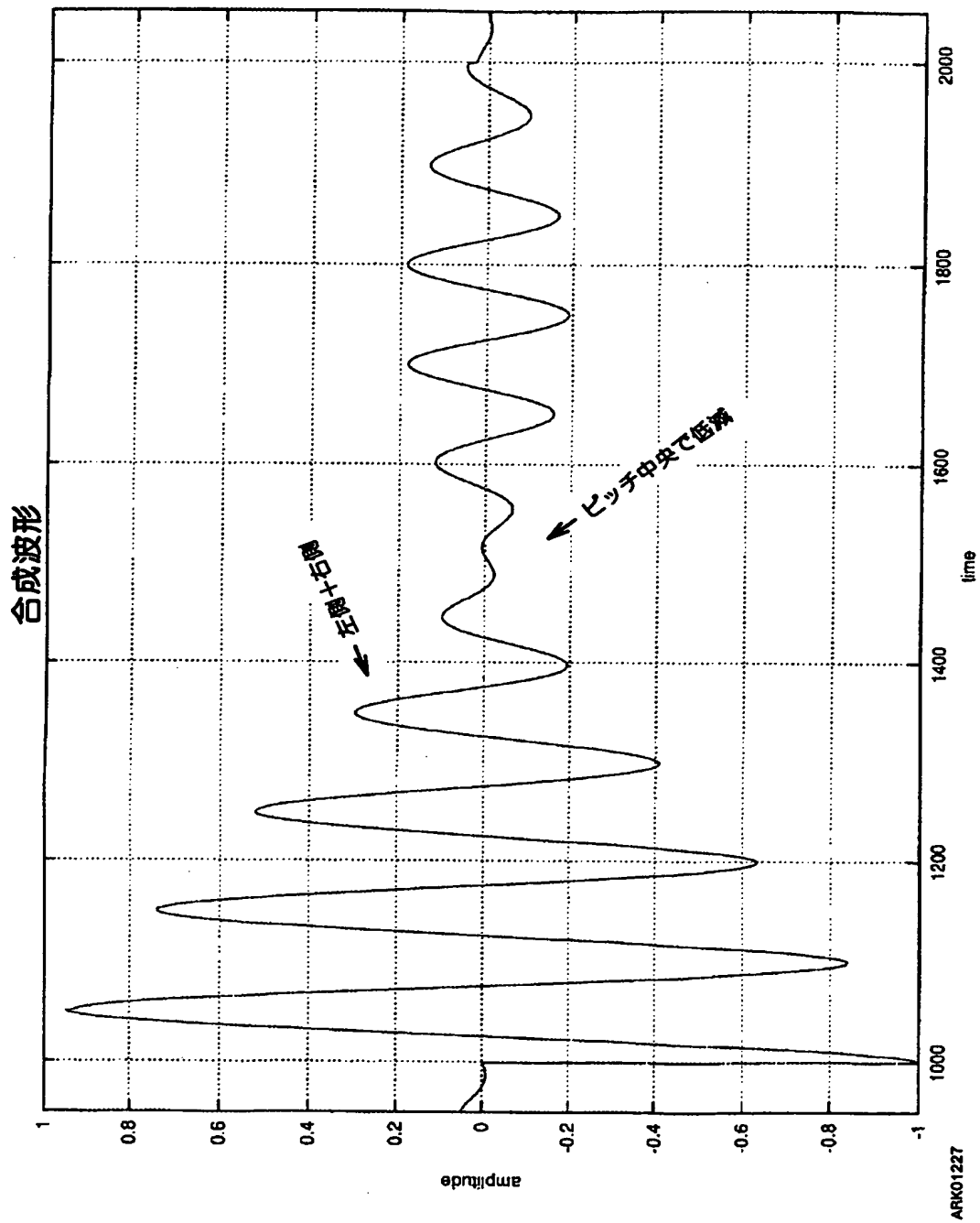
【図 26】

ARK012
図26



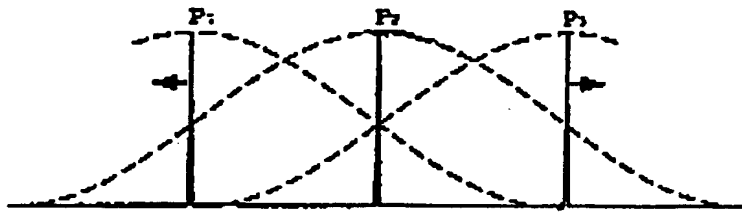
【図 2 7】

ARK012
図27

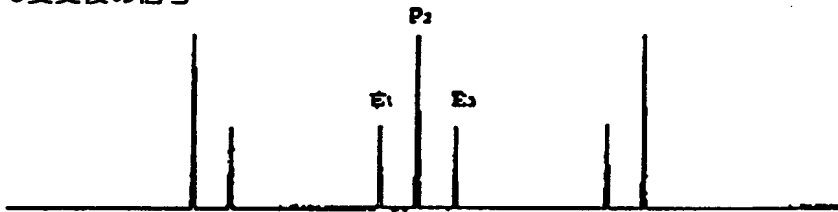


【図 28】

原信号（実線）および窓関数（破線）



F₀変更後の信号



ARK01228

【書類名】 要約書

【要約】

【課題】 音声の自然性を損なうことの少ないピッチ変更処理技術を提供する。

【解決手段】

音声波形の各ピッチにおける波形は、図 1 に示すように、1) 声帯の閉鎖に伴うマイナスピークが出現する時点 α 、2) マイナスピークを起点とした声道の形状に基づく波形が現れる区間 β 、3) 声道形状に基づく音声波形が減衰しながら次のマイナスピークに向かって収束する区間 γ に分けられる。発明者らは、上記の音声波形の特性に基づいて、声帯の閉鎖に伴うマイナスピークによる影響が小さい部分（つまり次のマイナスピークの直前である区間 γ ）において、ピッチ変更のための波形加工を行うことを着想した。これによれば、ピーク付近での波形形状を完全に維持しつつ、かつ、ピッチ変更の影響を最小限に抑えつつ、加工を行うことができる。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [593160323]

1. 変更年月日	1993年 8月30日
[変更理由]	新規登録
住 所	大阪府箕面市坊島1丁目3番40号
氏 名	株式会社アルカディア